AD-A191 605

2

DTIC FILE COPY

NAMRL- 1333

PREDICTING AIR COMBAT MANEUVERING (ACM) PERFORMANCE: FLEET FIGHTER ACM READINESS PROGRAM GRADES AS PERFORMANCE CRITERIA

G. R. Griffin, T. R. Morrison, T. L. Amerson,
and P. V. Hamilton

October 1987

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
PENSACOLA, FLORIDA

Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY Unclassified 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NAMRL-1333	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Aerospace Medical Research Laboratory 6b. OFFICE SYMBOL (If applicable) 03	7a. NAME OF MONITORING ORGANIZATION Naval Medical Research and Development Command
6c. ADDRESS (City, State, and ZIP Code) NAS, Pensacola, FL 32508-5700	7b. ADDRESS (City, State, and ZIP Code) NMC, NCR, Bethesda, MD 20814-5044
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63706N; PROJECT NO. M0096001; TASK NO. 001; WORK UNIT ACCESSION NO. 0151

11. TITLE (Include Security Classification)
(U) Predicting Air Combat Maneuvering (ACM) performance: Fleet fighter ACM readiness program grades as performance criteria

12. PERSONAL AUTHOR(S)
G. R. Griffin, T. R. Morrison, T. L. Amerson and P. V. Hamilton

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 1985 TO 1987	OCT 1987	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Pilots, Prediction, Performance, Criteria, Tactical Aircrew Combat Training System, Fleet Fighter ACM Readiness Program

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A difficult aspect of predicting fleet pilot performance is acquiring meaningful and reliable, inflight criteria. Without such criteria, performance assessment is both theoretically and realistically impossible. This study was an attempt to predict Air Combat Maneuvering (ACM) performance using performance-based laboratory tests and to evaluate the VF-43 adversary squadron's grading of inflight ACM performance in the Fleet Fighter ACM Readiness Program at Naval Air Station Oceana. The purpose of the latter effort was to select convenient and reliable criteria for ACM performance assessment and use in the validation of the laboratory tests.

In an initial evaluation (Study I), F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed performance-based perceptual motor and multitask tests. Results indicated that dichotic listening test measures, obtained during multitask conditions, could

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL J. O. HOUGHTON, CAPT MC USN, Commanding Officer	22b. TELEPHONE (Include Area Code) (904) 452-3286 — 22c. OFFICE SYMBOL 00

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

U.S. Government Printing Office: 1985—F07-047

19. ABSTRACT (Continued)

be used to reliably predict ACM inflight criteria. Results of a larger sample of F-14 pilots (Study II) indicated that an overall ACM grade (OAG) assigned by VF-43 adversary personnel can be predicted reliably by an objective kill difference composite score and three subjective measures: situational awareness, mutual support, and energy management. These four measures accounted for 78% of the variance with the OAG. A correlational analysis suggests that the VF-43 grading process is reliable and consistent.

Additional results were obtained on the relation between the Naval Aerospace Medical Research Laboratory vision tests and ACM criteria (Study III). Contrast sensitivity measures were significantly related to a mean time-to-first-kill measure. Visual acuity and accommodative flexibility measures were significantly related to the initial sighting (tally-ho) and visual identification (VID) of adversary aircraft on an instrumented range. Age and/or experience in ACM may be an important variable in relating vision tests to pilot performance.

It is recommended that: (1) improved performance-based tests should be administered to a sample of Navy pilots performing in Fleet Fighter ACM Readiness Evaluations to replicate initial test results; and (2) an overall ACM grade regression equation should be applied to a supplementary sample of pilots performing in Fleet Fighter ACM Readiness exercises to confirm the reliability and validity of the VF-43 adversary squadron's grading process.

SUMMARY PAGE

THE PROBLEM

A difficult aspect of predicting fleet pilot performance is acquiring meaningful and reliable, inflight criteria. Without such criteria, performance assessment is both theoretically and realistically impossible. This study was an attempt to predict Air Combat Maneuvering (ACM) performance using performance-based laboratory tests and to evaluate the VF-43 adversary squadron's grading of inflight ACM performance in the Fleet Fighter ACM Readiness Program at Naval Air Station Oceana. The purpose of the latter effort was to select convenient and reliable criteria for ACM performance assessment and use in the validation of the laboratory tests.

FINDINGS

In an initial evaluation (Study I), F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed performance-based perceptual motor and multitask tests. Results indicated that dichotic listening test measures, obtained during multitask conditions, could be used to reliably predict ACM inflight criteria. Results of a larger sample of F-14 pilots (Study II) indicated that an overall ACM grade (OAG) assigned by VF-43 adversary personnel can be predicted reliably by an objective kill difference composite score and three subjective measures: situational awareness, mutual support, and energy management. These four measures accounted for 78% of the variance with the OAG. A correlational analysis suggests that the VF-43 grading process is reliable and consistent.

Additional results were obtained on the relation between the Naval Aerospace Medical Research Laboratory vision tests and ACM criteria (Study III). Contrast sensitivity measures were significantly related to a mean time-to-first-kill measure. Visual acuity and accommodative flexibility measures were significantly related to the initial sighting (tally-ho) and visual identification (VID) of adversary aircraft on an instrumented range. Age and/or experience in ACM may be an important variable in relating vision tests to pilot performance.

RECOMMENDATIONS

1. Improved performance-based tests should be administered to a sample of Navy pilots performing in Fleet Fighter ACM Readiness Evaluations to replicate initial test results.

2. An overall ACM grade regression equation should be applied to a supplementary sample of pilots performing in Fleet Fighter ACM Readiness exercises to confirm the reliability and validity of the VF-43 adversary squadron's grading process.

Acknowledgments

The authors gratefully acknowledge Mr. Al Thomas and HM3 Arelia Walking Eagle for administering multitask tests and gathering TACTS ACM performance data at the Naval Air Station Oceana, Virginia Beach, Virginia, and Dr. Ailene Morris and Mr. Ed Ricks for administering vision tests and collecting vision test data.

This research was sponsored by the Naval Medical Research and Development Command and was performed under Work Unit No. 63706N M0096.001-0151.

INTRODUCTION

Research is ongoing at the Naval Aerospace Medical Research Laboratory to predict fleet aviator inflight performance using perceptual psychomotor and information processing tasks. The goal is to develop relevant laboratory tasks, test aircrew, and relate aircrew test performance to simulated and actual flight performance. From this effort, it may be possible to aid decisions concerning aircrew selection, training pipeline assignment, and post-training aircraft assignment. Crucial to this research is the identification of useful, valid, and reliable measures of flight performance for the validity assessment of predictor tests.

Previous United States Navy research to predict operational performance has been encouraging (1-7). Rickus and Berkshire (4) reported that peer ratings obtained in Navy preflight training were useful in identifying successful and unsuccessful aviators in combat (Viet Nam). Bale et al. (2) evaluated F-4 Replacement Air Group (RAG) training during the mid-60s and developed a prediction equation that could reduce RAG attrition from 13.3 to 8.3%. A study of Tactical Aircrew Combat Training System (TACTS) F-4 air combat maneuvering by Ciavarelli et al. (1) in the late 70s found three measures (angle-off-tail, closing velocity, and indicated air speed) that were significantly related to ACM performance. Bricxson et al. (3) were able to successfully predict F-4 carrier landing performance. Shannon et al. (7) found that a relatively small set of RAG measures can reliably predict final overall RAG grade (multiple $\underline{R}$ = .84). The two most important measures (carrier qualification power/nose control and offensive ACM) accounted for 73% of the variance with the final overall RAG grade. In two subsequent studies, Shannon and Waag (6) found that an equation based on an east coast RAG reliably predicted performance of F-4 pilots on a west coast RAG and reported (5) that experience and seven undergraduate training grades reliably predicted final overall RAG grade (multiple $\underline{R}$ = .51).

Despite these positive results, new aircraft and weapons system technology may have made the research obsolete. In addition, the approaches of the previous studies and the present effort differ. Previous studies (2-6) used pencil-and-paper selection tests and undergraduate training measures to predict performance. Our approach used performance-based tests of cognitive, perceptual, and multitask functioning to predict fleet operational aviator performance.

The present study represents an attempt to predict ACM inflight performance using performance-based automated tests, and an evaluation of the VF-43 adversary squadron's grading of aircrew performance in the Fleet Fighter ACM Readiness Program at Naval Air Station Oceana. The purpose of the latter effort was to select useful and reliable criteria for ACM performance assessment and validating future laboratory tests. In addition, correlations between measures of the TACTS and vision tests were determined.

Study I. Multitask Test and Marine Pilot ACM Performance

The purpose of this evaluation was to test the feasibility of predicting ACM performance with performance-based perceptual-motor and cognitive multitask tests.

SUBJECTS

Twenty-two F-4 pilots from Marine Squadron 451 served as subjects during their participation in a Fleet Fighter ACM Readiness Program exercise at NAS Oceana during the summer of 1985.

PROCEDURE

Air combat maneuvering performance data are routinely collected at the NAS Oceana TACTS facility. The data are used by VF-43 adversary squadron personnel to develop aircrew and squadron ACM performance ratings. The performance ratings provide training feedback to individual aircrews by highlighting their strengths and weaknesses in ACM, and also provide a method for military managers to assess overall squadron readiness. In addition, the TACTS Fleet Fighter ACM Readiness Exercise results serve as a base to evaluate the tactical employment of aircraft and weapon systems. Typical data resulting from the readiness exercises are presented in Table 1. A description of the TACTS training system and definitions for specific TACTS performance measures are in Appendixes A and B, respectively.

Eighteen of the 22 Marine pilots completed single- and multitask cognitive and perceptual-motor tests during the readiness exercise. The tests consisted of a 24-trial dichotic listening task (DLT) followed by 6-min performance on a psychomotor task. Both tasks were then performed simultaneously. Correlational and multiple regression analyses were conducted on the ACM performance measures to identify suitable criteria and to evaluate the strengths of the correlations among the various measures. Subjects' test performance data were correlated with the identified criteria.

RESULTS

Pearson correlations of 27 measures associated with the VF-43 adversary squadron's evaluation of ACM performance of 22 F-4 pilots are presented in Table 1. The overall ACM grade (OAG) was significantly related to offensive maneuvering (r = .67), situational awareness (r = .81), and mutual support (r = .56). In addition, the OAG was significantly related to mean time to first kill (r = -.42), number of VF-43 adversary squadron missile shots (r = -.65), and the number of times a pilot was "killed" (r = -.70) in the simulated exercises. The negative correlations indicated that a higher ACM grade was associated with shorter times to first kill, fewer adversary squadron missile shots taken, and fewer times being "killed" in the simulated exercises (better ACM performance).

A multiple regression analysis indicated that situational awareness, offensive maneuvering, number of times killed, and mutual support accounted for 89% of the variance associated with the OAG criterion (R = .95, $F(4, 17)$ = 36.19, p < .0001). The OAG and these four measures were then correlated with the single- and multitask cognitive and psychomotor test performance of 18 of the 22 pilots[1]. A derived composite kill-difference score (the total number of ACM kills minus the number of times a pilot was killed in the TACTS simulated exercises) was included in the correlational analyses as

[1] Four pilots did not volunteer to complete the series of tests.

well as the total number of flight hours, which ranged from 337 to 1925. Of the 42 correlations computed between the tests and ACM criteria, 4 (10%) were significant at the .05 or .01 level of confidence.

TABLE 1. Pearson Correlations of Individual ACM Performance Measures with Overall ACM Grade(a) (N = 22).

Subjective measures	r
Use of environment	.05
Start/VID start	.07
First move	-.18
Aggressiveness	.16
Offensive maneuvering	.67**
Defensive maneuvering	-.04
Keeping sight/lookout	.39
Energy management	.23
Mental plot	-.17
Situational awareness	.81**
Bugout technique	-.13
Weapon system employment	.24
VID technique	-.08
VID communication	.05
UHF communication	.24
Game plan usage	.16
Mutual support	.56**
Debrief	-.17
Reconstruction	-.37
Objective Measures	r
Total number TACTS kills	.31
Number of missiles launched	.16
Mean time (s) to first kill	-.42*
Visual tally-ho mean range	.01
VID mean range	-.05
Number of times killed	-.70**
Number VF-43 missile shots	-.65**

* p < .05
** p < .01

(a) The overall grade is a composite of the 19 subjective measures.

Note: Performance measure definitions are in appendix B.

The Pearson correlations of the cognitive and psychomotor tests shown in Table 2 indicated that a DLT measure (DLT-1) obtained during multitask performance was significantly related to offensive maneuvering (r = .62) and the kill-difference composite score (r = .49). A DLT multitask measure

based on a slightly different scoring procedure (DLT-2) was significantly related to the OAG ($\underline{r}$ = .49) and the offensive maneuvering score ($\underline{r}$ = .60). Number of flight hours was unrelated to any ACM or test performance measure. These results, although based on a small sample of Marine pilots, support the feasibility of developing a series of performance-based cognitive, perceptual, and multitask tests to predict aviator performance.

TABLE 2. Correlation of ACM Performance Criteria and Single and Multitask Cognitive and Psychomotor Tests ($\underline{n}$ = 18 F-4 Pilots).

Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. Overall ACM grade												
2. Situational awareness	.83**											
3. Offensive maneuvering	.83**	.61**										
4. Number times killed	-.75**	-.58*	-.58*									
5. Mutual support	.51*	.58*	.25	-.21								
6. Kill-difference score	.50*	.25	.41	-.59**	-.03							
7. Flight hours	.13	-.01	.06	-.39	.21	.05						
8. Multitask DLT 1	.43	.10	.62**	-.27	-.16	.49*	.11					
9. Multitask DLT 2	.49*	.12	.60**	-.33	-.03	.37	.27	.93**				
10. Multitask RMT	-.14	.04	-.25	.13	.20	-.05	-.18	-.40	-.45			
11. Single task DLT 1	.21	-.02	.28	-.17	-.15	.35	.04	.56*	.53*	-.72**		
12. Single task DLT 2	.32	.15	.36	-.20	-.19	.22	-.10	.60**	.55**	-.67**	.85**	
13. Single task RMT	.10	-.03	.16	-.18	-.29	.16	.07	.27	.29	.31	-.12	-.06

*.05 = .47
**.01 = .59

Unresolved was a series of important questions concerning the Fleet Fighter ACM Readiness Program evaluation process: Are the resulting grades reliable for Navy pilots flying contemporary F-14 aircraft? What is the relation of the VF-43 grading process to more objective TACTS ACM performance measures (i.e., total number of kills, visual identification (VID) range, VID kills and engaged kills)? Which ACM measures are most predictive of ACM performance?

Study II: Fleet Fighter ACM Readiness Program Grades as Criteria

The purpose of the second study was to answer the questions posed above, and assess the utility of VF-43 ACM grades and TACTS ACM performance measures as criteria for the validation of tests developed to predict ACM performance.

SUBJECTS

Subjects were 125 Navy F-14 pilots (10 fighter squadrons) who participated in the Fleet Fighter ACM Readiness Program against the VF-43 adversary squadron at NAS Oceana during 1985 and January of 1986.

PROCEDURE

Air combat maneuvering "competitive exercise" performance data were collected for the Navy participants in the Fleet Fighter ACM Readiness Program and analyzed to derive correlational statistics. Multiple regression analyses were performed to study the relative importance of specific predictors and to derive criteria that would predict the overall ACM grade.

RESULTS

Performance measure descriptive statistics and Pearson correlations between the OAG and 19 subjective and 12 objective TACTS measures are presented in Table 3. The kill-difference composite score was the measure most highly correlated with OAG (r = .76), followed by the engaged kill-difference composite score (r = .70). The total number-of-kills measure was related to OAG (r = .65), as were missiles launched (r = .58), VID-kill (r = .39) and engaged-kill (r = .57) scores. As expected, the number-of-times-killed measure was significantly and negatively related to OAG (r = -.51). Number of radar locks was significantly correlated with the OAG (r = .24) as well. Four of the objective measures--mean time-to-first-kill, radar lock mean range, visual tally-ho mean range, and VID mean range measures--were not significantly related to OAG.

We were surprised that the mean time-to-first-kill score was unrelated to OAG. One explanation for this result may be that the time-to-first-kill score is an average of both VID and engaged-kill times. This pooling of relatively short (VID) and longer (engaged) kill times may have a confounding effect on the resulting correlations. Separation of VID and engaged kill times might enable a better understanding of the relation of this ACM score with the OAG and other TACTS measures.

An examination of the subjective measures, as shown in Table 3, indicated that 11 measures were significantly correlated with the OAG. Situational awareness (described by VF-43 adversary personnel as a synonym for ACM proficiency) correlated most highly with OAG (r = .70), followed by offensive maneuvering (r = .53), aggressiveness (r = .45), mutual support (r = .44), and start/VID start (r = .40). Defensive maneuvering, keeping sight, energy management, weapon system employment, VID technique, and game plan measures were significantly correlated with the OAG as well, with correlations between .23 and .39. Those measures not significantly correlated with the OAG were use of environment, first move, mental plot, bugout technique, VID communication, UHF communication, debrief, and reconstruction.

The full correlation matrix is presented in Appendix C. Of 496 correlations, 163 (33%) were significant at the .05 level or above.

TABLE 3. Performance Measure Descriptive Statistics and Pearson Correlations Between Overall ACM Grade and ACM Performance Measures (N = 125).

Subjective measures	r	Mean	SD	Min	Max
Overall grade	--	2.01	.05	1.90	2.15
Use of environment	.00	2.01	.05	1.83	2.20
Start/VID start	.40**	2.04	.16	1.67	2.88
First move (n = 113)	.12	2.01	.18	1.50	2.50
Aggressiveness	.45**	2.08	.13	1.88	2.75
Offensive maneuvering	.53**	2.08	.18	1.67	2.50
Defensive maneuvering	.39**	1.96	.14	1.56	2.25
Keeping sight/lookout	.35**	1.95	.14	1.50	2.29
Energy management	.37**	2.00	.15	1.57	2.38
Mental plot	-.11	1.98	.11	1.50	2.90
Situational awareness	.70**	1.93	.25	1.25	2.75
Bugout technique	.11	2.03	.17	1.67	2.50
Weapon system employment	.30**	2.06	.19	1.25	2.50
VID technique	.30**	1.99	.12	1.67	2.33
VID communication	-.08	1.99	.05	1.75	2.17
UHF communication	.12	2.00	.08	1.50	2.25
Game plan usage	.32**	2.08	.18	1.50	2.50
Mutual support	.44**	1.95	.26	1.33	2.50
Debrief (n = 112)	.11	2.01	.07	2.00	2.50
Reconstruction	.16	2.01	.04	2.00	2.25
Objective measures	r	**Mean**	**SD**	**Min**	**Max**
Total number TACTS kills	.65**	5.63	2.80	0.0	14.0
Number of missiles launched	.58**	13.14	6.85	1.0	31.0
Number of VID kills	.39**	2.66	1.64	0.0	7.0
Number of engaged kills	.57**	2.97	2.05	0.0	10.0
Mean time (s) to first kill (n =124)	-.02	42.98	25.31	6.0	115.3
Number of radar locks	.24**	4.95	1.55	0.0	7.0
Radar locks mean range	-.04	13.82	3.22	0.0	26.0
Visual tally-ho mean range	.12	2.80	1.37	0.0	6.5
VID mean range	.16	1.59	0.88	0.0	5.2
Number of times killed	-.51**	1.68	1.10	0.0	5.0
Kill-difference score	.76**	3.92	3.14	-2.0	13.0
Engaged-kill-difference score	.70**	1.29	2.44	-3.0	9.0

** $p < .01$

REGRESSION ANALYSIS

To examine which subjective and objective measures would best predict the OAG, a series of forward selection multiple regression analyses (8) was conducted. A forward selection stepwise multiple regression technique was used because of multi-colinearity (high intercorrelations) among certain of the objective and subjective measures. The first regression (Appendix D, Table D-1) was based on the subjective measures in Table 3. Results indicated that a 6-measure regression model accounted for 83% of the

variance with OAG ($R = .91$, $F(6, 113) = 98.57$, $p < .0001$). The situational awareness measure entered the regression first and accounted for 49% of the variance with the OAG. Offensive maneuvering, mutual support, start/VID start, energy management, and keeping sight measures then entered the regression equation accounting for 11, 7, 8, 4, and 4% additional variance, respectively.

A second regression analysis (Appendix D, Table D-2) was conducted using the objective performance measures in Table 3. The measures total TACTS kills, kill-difference score, and engaged kill-difference score were excluded because they represented combinations of other measures. As shown in Appendix C, number of missiles launched was related to total TACTS kills ($r = .83$), engaged-kills ($r = .67$), VID-kills ($r = .57$), and the kill-difference score ($r = .78$). Since this measure is simply a means of achieving TACTS kills, it too was excluded. These composite and/or duplicative measures were omitted from the regression to increase insight into those specific performance measures most important to the OAG. The results of the multiple regression indicated that engaged-kills, number-of-times-killed, and VID-kills accounted for 62% of the variance with the OAG. The engaged-kill measure entered the model first and accounted for 33% of the variance associated with the OAG. Number-of-times-killed and VID-kill measures followed in succession, accounting for 19 and 10% additional variance, respectively. Finally, the mean time-to-first-kill measure entered the regression model but accounted for only 1% additional variance ($R = .79$, $F(4, 120) = 51.02$, $p < .0001$). It is important that both engaged-kills and VID-kills entered the regression model (both are significantly correlated with OAG, but the correlation between the two measures is low, $r = .14$). These results suggest that the VID-kill and engaged-kill performance measures are statistically independent in this population of Navy pilots. They also emphasize the importance of pilot training in both of these ACM skills.

A third multiple regression model (Appendix D, Table D-3) was based on a kill-difference score (a composite of the first three measures entering the second regression model) and the subjective measures of Table 3. The kill-difference measure entered the regression first, accounting for 57% of the variance with OAG. Next, the situational awareness measure entered the regression, accounting for an additional 14% of variance, followed by energy management and mutual support, which each contributed about 4% additional variance ($R = .89$, $F(4, 120) = 109.39$, $p < .0001$).

These results, indicating the importance of kills in the VF-43 adversary squadron's grading of ACM performance, were expected since kill ratios from the competitive exercises of the Fleet Fighter Readiness Program represent a basic component of the grading process (9). Unexpectedly, situational awareness and other subjective measures contribute important additional variance in the prediction of OAG. Apparently human judgement of ACM proficiency is an important element in these performance evaluations.

SITUATIONAL AWARENESS

At NAS Oceana, VF-43 adversary personnel define situational awareness as the "total of ACM." This definition seems appropriate based on the results reported here. Table 3 shows that situational awareness is the subjective measure most strongly related to OAG ($r = .70$). Item 11 in Appendix C shows

the various correlations between situational awareness and the other objective and subjective measures.

Those measures most strongly related to situational awareness, in addition to the OAG, are the kill-difference score, engaged-kill-difference score, number-of-times-killed, VID-kills, engaged-kills, total TACTS kills, and number of missiles launched. Subjective measures--start/VID start, aggressiveness, offensive maneuvering, defensive maneuvering, keeping sight, VID technique, game plan and mutual support--are also significantly related to the situational awareness measure.

Objective measures unrelated to situational awareness (in this analysis) are visual tally and VID range, number of radar locks, radar lock range, and mean time-to-first-kill. Subjective measures that are not significantly related to situational awareness are environment, first move, energy management, mental plot, bugout, weapon employment, VID and UHF communication, debrief, and reconstruction.

PERFORMANCE MEASURE RELIABILITY

An important aspect of this study concerns the reliability or consistency of the VF-43 performance measures. To evaluate the reliability of the TACTS objective performance scores and the more subjective VF-43 scoring process, we randomly divided the Navy pilot sample in half and examined performance measure correlations with the OAG (Table 4). Subjects were divided on the basis of even/uneven chronological subject number[2]. Table 4 includes Pearson correlations based on the total sample to allow comparison with the correlations of both subsamples. In addition, the absolute difference of the Pearson correlations are presented. Table 4 reveals remarkably similar results, especially for the more objective TACTS parameters. The one objective measure that indicated a major correlation change was the mean time-to-first-kill measure, with an r of .09 for the even and -.12 for the uneven subsample, an absolute difference of .21. This change in correlational value is not significant at the .05 level. Moreover, the mean time-to-first-kill measure is not significantly related to the OAG. All other objective measure correlational values were highly similar. Six of the subjective measures had an absolute difference correlation value of .20 or greater. Only two of these measures, UHF communication and reconstruction, represented a significant difference between the even and uneven pilot subsamples ($p < .05$), based on a Fishers Z test of significant differences of correlations. Neither of these correlational values, however, were significantly correlated with OAG for the total sample or the two subsamples. In summary, the objective and subjective measures most highly correlated with OAG differ only slightly for the two randomly derived samples.

A second approach to establishing the reliability of the OAG was to apply the regression model of Appendix D, Table D-3 (based on the kill-difference, situational awareness, energy management, and mutual support measures) to various pilot subsamples. This particular regression model was used because it represents the best prediction of OAG using both objective and subjective performance measures. A Pearson correlation value was

[2] Subject performance data were ordered for statistical analysis by date of the ACM readiness evaluation and the alphabetical order of pilot name.

computed between the predicted and actual OAG grade for eight different pilot subsamples (Table 5). Based on a Fishers Z transformation, the average of the eight correlation values is .88.

TABLE 4. Pearson Correlations for Total, Even, and Uneven Ordered Pilots, and Correlation Absolute Difference Scores.

Performance measure correlation with OAG	All pilots (N = 125)	Even (n = 62)	Uneven (n = 63)	Difference
Use of environment	.00	.03	-.03	.06
Start/VID start	.40**	.38**	.42**	.04
First move	.12	.05 (n=56)	.20 (n=57)	.15
Aggressiveness	.45**	.41**	.47**	.06
Offensive maneuvering	.53**	.49**	.57**	.08
Defensive maneuvering	.39**	.38**	.41**	.03
Keeping sight/lookout	.35**	.45**	.26*	.19
Energy management	.37**	.46**	.28*	.18
Mental plot	-.11	-.13	-.06	.07
Situational awareness	.70**	.63**	.76**	.13
Bugout technique	.11	.22	.01	.21
Weapon system employment	.30**	.39**	.18	.21
VID technique	.30**	.29*	.34**	.05
VID communication	-.08	.02	-.16	.18
UHF communication	.12	-.15	.33**	.48**(1)
Game plan usage	.32**	.42**	.22	.20
Mutual support	.44**	.38**	.49**	.11
Debrief	.11	.10 (n=55)	.12 (n=57)	.02
Reconstruction	.16	.32**	-.10	.42*(1)
Total number TACTS kills	.65**	.68**	.61**	.07
Number missiles launched	.58**	.60**	.56**	.05
Number of VID kills	.39**	.33**	.45**	.12
Number of engaged kills	.57**	.63**	.51**	.12
Mean time-to-first-kill	-.02	.09 (n=61)	-.12 (n-63)	.21
Number of radar locks	.24**	.23	.26*	.03
Radar locks mean range	-.04	.02	-.08	.10
Visual tally-ho mean range	.12	.04	.17	.13
VID mean range	.16	.07	.22	.15
Number of times killed	-.51**	-.47**	-.56**	.09
Kill-difference score	.76**	.79**	.73**	.06
Engaged kill-difference score	.70**	.74**	.67**	.07

\- - - - - - -

* $p < .05$
** $p < .01$
(1) Fisher Z test of significance of Pearson correlations

TABLE 5. Pearson Correlation Values for Eight Pilot Subsamples Based on Predicted and Actual OAG.

Pilot subsamples	*r*	Number
Even	.89	62
Uneven	.88	63
First half	.84	62
Second half	.90	63
1st, 3rd quarter	.88	62
2nd, 4th quarter	.89	63
2nd, 3rd quarter	.89	62
1st, 4th quarter	.89	63

Average *r* (based on Fisher Z transformations) = .88
All values significant, $p < .01$.

In summary, regardless of the sampling procedure, the model for predicting the OAG provided similar results. Because the grading of the aircrews by different adversary pilots seems consistent, we can assume that the internal criteria by which the grades are assigned are similar across adversary pilots. In essence, the grading process appears reliable.

PILOTS, AIRCREW, AND VISUAL PERFORMANCE

Although the F-14 aircraft normally employs both a pilot and a Radar Intercept Officer (RIO) working as a team, we addressed those measures associated with pilot performance. The VF-43 scoring process emphasizes pilot proficiency, since the pilot maneuvers the aircraft, fires the missiles, and as the aircraft commander is responsible for engagement outcome. However, it is important to realize that the RIO's efforts in operating the radar, keeping sight, and performing lookout have an important effect on ACM engagement outcome. Consequently, we included two RIO measures (number of radar locks and radar lock mean range) in this analysis to examine the relation of RIO performance to pilot tally-ho and pilot aircraft identification range (important in pilot visual performance). The importance of the RIO's radar skills and pilot visual performance is demonstrated by the significant Pearson correlations between the number of radar locks, visual tally-ho mean range, VID mean range, and other objective TACTS ACM scores of Table 6.

Initially, it was unclear as to why an RIO performance parameter (number of radar locks) would be significantly related to pilot visual tally-ho ($r = .43$) and VID mean range ($r = .50$). After careful consideration, we believe that a radar lock acts to decrease the visual target search area for the pilot, who then can attend to the diamond[3], knowing that an adversary aircraft will ultimately become a visual target at the indicated location on the head-up display.

[3] An area of the head-up display, delineated as a diamond shape, indicating the location of the radar target.

TABLE 6. Pearson Correlations Between Radar Locks, Pilot Visual Tally and VID, and Objective ACM Performance ($\underline{N}$ = 125 Navy Pilots).

ACM objective measure	Number of radar locks	Radar lock mean range	Visual tally mean range	VID mean range
Total number TACTS kills	.41**	-.10	.30**	.41**
Number missiles launched	.40**	-.21*	.32**	.42**
Number of VID kills	.42**	.12	.45**	.48**
Number of engaged kills	.22*	-.05	.05	.16
Mean time-to-first-kill	-.12	-.01	-.19*	-.14
Number of radar locks	-	.25**	.43**	.49**
Radar lock mean range	.25**	-	.05	-.07
Visual tally-ho mean range	.43**	.05	-	
VID mean range	.50**	-.07	.68**	-
Number of times killed	.06	.00	.10	.07
Kill-difference score	.34**	-.09	.23**	.33**
Engaged-kill-difference score	.14	-.05	.00	.11

* $\underline{p} < .05$
** $\underline{p} < .01$

A radar lock is important to kills, and it is a requirement for successful use of a forward aspect missile. We found significant correlations between number of radar locks and VID kills ($\underline{r}$ = .42) and, to a lesser extent, engaged kills ($\underline{r}$ = .22). The number of radar locks was significantly related to total number of TACTS kills ($\underline{r}$ = .41), number of missiles launched ($\underline{r}$ = .40), and the kill-difference score ($\underline{r}$ = .34). Radar lock mean range was not a significant predictor for the majority of objective ACM performance measures. Apparently, when radar lock is accomplished, it occurs at distances so great that the variability in lock ranges does not influence subsequent ACM performance. Failure to acquire radar lock is another matter, however, as noted above.

Visual tally-ho mean range and VID mean range, as previously noted, are strongly correlated with the radar lock measure. Apparently, a radar lock significantly enhances the pilot's acquisition of visual targets. Since VID of adversary aircraft is required before missile launch, under present tactical rules, vision-dependent ACM performance measures should be positively related to number of TACTS kills. Our data support this hypothesis. That is, a greater visual tally-ho range and greater VID range are each significantly associated with a greater number of TACTS kills ($\underline{r}$ = .30 and .41, respectively). Further, we hypothesized that vision-dependent ACM performance might be more highly related to the number of VID kills rather than engaged kills[4]. Our data support this hypothesis also. Visual tally-ho performance

[4] VID kills are those that occur immediately following initial target detection and identification and are generally made with radar directed missiles fired head-on. Engaged kills occur during subsequent dogfighting, when pilots attempt to maneuver behind their adversary to fire guns or heat seeking missiles.

is significantly correlated with VID kills (*r* = .45) but not engaged kills (*r* = .05). Visual identification performance is also significantly related to VID kills (*r* = .48) and not engaged kills (*r* = .16).

Number of radar locks, visual tally-ho mean range, and VID mean range are negatively related to the mean time-to-first-kill measure. These correlations are negative since the launch of a forward aspect missile (the best means of achieving a quick kill) generally depends on achieving each of these measures in sequence. Having a longer tally-ho or VID range enables better aircrew preparation at the merge and reduces time-to-first-kill. Additionally, a radar lock allows more certainty in visual search and produces longer range visual target acquisition and aircraft identification. In summary, the relation of radar locks to improved vision-dependent ACM performance and the relation of visual tally-ho and VID performance to subsequent missile launch and a VID kill reflect a necessary sequence of performance events for achieving mission success.

Study III: The Relation of Vision Test, Experience, and ACM Performance

A series of psychophysical vision tests developed at the Naval Aerospace Medical Research Laboratory (NAMRL) is currently being evaluated to determine critical visual skills required in naval aviation. The availability of both TACTS ACM performance and vision test data represented an opportunity to study the correlational relation between the two. In addition, pilot age and flight hours, collected as a part of the vision test effort, enabled us to study the relation of experience with TACTS ACM performance.

SUBJECTS

Eighty-nine of the 125 Navy pilots of this evaluation participated in a visual testing evaluation at NAS Oceana. Not all subjects completed all tests in the vision test battery.

PROCEDURE

Relevant ACM criteria (OAG, situational awareness, kill-difference score, VID-kills, engaged-kills, and number of times killed) of Study II were correlated with vision test scores. Additionally, TACTS visual tally-ho, and VID range scores were included because of their relevance to pilot vision test performance. The mean time-to-first-kill measure was included because it might be related to pilot experience. Pilot experience measures were age and flight hours (jet hours, total flight hours, TACTS flight hours, and total ACM hours) as reported by subjects during vision testing.

RESULTS

Table 7 presents the Pearson correlations for the vision test and ACM performance criteria. One vision test measure, spot detection threshold stress response time (SPOT-SRT), was significantly related to the OAG (*r* = -.22). The negative correlation indicates that longer spot detection response times (i.e., slower performance) were associated with lower (poorer) OAGs. Vision test scores were not significantly related to situational awareness, VID kill, or engaged kill criteria. Contrast sensitivity measures at spatial frequencies (CS 3.0 cycles/degree, *r* = -.24;

CS 11.4 cycles/degree, $r = -.26$; and CS 22.8 cycles/degree, $r = -.25$) were significantly related to the mean time-to-first-kill criterion. In addition, a high contrast acuity threshold mean measure (ACHI-TM) was significantly related to mean time to first kill ($r = -.21$). Generally, for the five contrast sensitivity spatial frequencies and the high contrast acuity threshold test, poorer vision test scores were associated with shorter mean times to first kills (i.e., better ACM performance). The consistent direction and magnitude of the correlations between contrast sensitivity and mean time to first kill indicate a reliable effect, although not in the expected direction. We expected better vision test scores to be associated with better ACM performance, however, mean time to first kill was inversely related to age, jet hours, total flight hours, and TACTS hours, as shown in Table 8. Older and more experienced pilots achieved shorter mean times to first kill. The literature indicates a progressive general deterioration in visual functions with age. In particular, visual acuity and contrast sensitivity decrease significantly with age (10,11).

TABLE 7. Pearson Correlations for the Vision Test and ACM Performance Criteria (n varies).

Vision test	n	OAG	Situational awareness	Mean time to first kill	Visual tally mean range	VID mean range	Number times killed	VID kills	Engaged kills	Kill difference score
ACHI-TM	(89)	.05	.15	-.21*	-.06	-.04	-.09	.14	-.10	.00
ACHI-SRT	(88)	-.08	.10	.04	.19	.15	.12	-.09	.04	-.C0
ACLO-TM	(89)	.07	.12	-.13	-.22*	-.07	-.13	-.00	-.09	-.01
ACLO-SRT	(88)	-.04	.05	.08	.21*	.18	.05	.01	.18	.11
GLAR-TM	(89)	-.02	.04	-.13	-.26**	-.12	-.22*	.03	-.15	-.08
GLAR-SRT	(88)	-.15	-.13	.07	.12	.05	.07	.03	.12	.08
GLRV-TM	(61)	.16	.13	-.12	-.14	-.03	-.24	-.00	.07	.06
GLRV-SRT	(61)	-.05	-.03	.11	.26*	.34**	.10	-.07	-.05	.03
SPOT-TM	(89)	-.06	.08	-.16	-.22*	-.03	-.06	.10	-.09	-.12
SPOT-SRT	(88)	-.22*	-.09	.02	.12	.03	.12	.10	-.07	-.05
FXFN-TM	(63)	-.12	-.07	-.08	-.02	.03	-.08	.08	-.07	-.17
FXFN-SRT	(63)	.01	-.09	-.11	.25*	.37**	.16	.13	-.05	.10
FXNF-TM	(63)	.01	.21	-.23	-.05	-.13	-.05	-.08	-.18	-.16
FXNF-SRT	(63)	-.04	-.08	.05	.13	.35**	.06	-.07	-.06	.07
CS 0.5	(71)	.04	.02	-.22	.12	.24*	.12	.00	-.03	.01
CS 1.0	(71)	-.09	-.09	-.17	.14	.19	.13	-.01	-.16	-.15
CS 3.0	(71)	-.08	-.03	-.24*	.08	.32**	.04	-.03	-.18	-.14
CS 6.0	(71)	-.06	-.05	-.22	.05	.11	.02	.03	-.18	-.15
CS 11.4	(71)	-.09	-.01	-.26*	.04	.12	.01	.04	-.22	-.12
CS 22.8	(71)	-.08	-.12	-.25*	-.03	-.06	-.04	.11	-.21	-.24*

* $p < .05$
** $p < .01$

Note: See text for vision test identification.

Our results suggest that experience is mediating the inverse relation between contrast sensitivity and acuity, and mean time to first kill. That is, pilots who obtained shorter mean times to first kills had poorer contrast sensitivity and acuity. They also were more experienced (and older) and apparently used this experience to achieve faster kills in ACM.

A contrast sensitivity measure (CS 22.8 cycles/degree) was significantly related to the kill-difference score (r = - .24). The contrast sensitivity test score is a negative number score. A greater negative number indicates better performance. Thus, a negative correlation means that better contrast sensitivity is associated with a greater kill-difference score. The direction of the relationship between contrast sensitivity and the kill-difference score is opposite that between contrast sensitivity and mean time to first kill. We have indicated previously that age may be a possible mediating variable between contrast sensitivity and mean time to first kill, and may account for the finding that poor contrast sensitivity is associated with better ACM performance. Now, results are presented that do not support this interpretation, i.e., better contrast sensitivity was associated with an improved kill-difference score. However, as shown in Table 8, the age and experience variables were related to mean time to kill but generally not to the kill-difference score. Thus, our data may not support an experience effect mediating the relationship between contrast sensitivity and the kill-difference score, but does regarding mean time to first kill.

As expected, the vision test scores were related more to the highly vision-dependent components of ACM, e.g., visual tally range and VID range than to other criteria. Both low contrast acuity threshold mean (ACLO-TM) and low contrast acuity threshold stressed response time mean (ACLO-SRT) correlated significantly with visual tally range (r = -.22 and r = .21, respectively). Low contrast acuity with glare threshold mean (GLAR-TM) correlated significantly with visual tally range as well (r = -.26). Spot detection-threshold mean (SPOT-TM) (r = -.22) also correlated significantly with visual tally range.

Low contrast acuity (with glare and visor) threshold stress response time mean (GLRV-SRT) was significantly correlated with both visual tally range (r = .26, and VID range (r = .34). Also, an accommodative flexibility (far to near) threshold stress response time measure (FXFN-SRT) was significantly correlated with both visual tally range (r = .25), and VID range (r = .37). The accommodative flexibility (near to far) threshold stress response time measure (FXNF-SRT) was significantly correlated with VID range (r = .35) but not visual tally range. In every case, the significant correlations between vision tests and visual tally/VID range were in the expected direction. For each of the significant correlations reported above, the threshold mean (TM) values are negatively correlated with visual tally and VID range, and the threshold stress response time (SRT) measures were positively correlated with the visual tally or VID range measures. These inverse relationships between the TM and SRT measures have been reported elsewhere (12,13) and seem to be reliable effects. Apparently, aviators with a low threshold response mean score (better vision performance) achieve visual tally and VID at longer ranges. They also have longer threshold stress response times, thus producing negative correlations between SRT measures and visual tally and VID ranges. It is uncertain why this occurs. One hypothesis is that better performers on the threshold mean vision tests are more deliberate, and hence slower, in making responses.

Other visual tests that correlated significantly with VID range (but not visual tally range) were contrast sensitivity at 3.0 cycles/degree (*r* = .32) and 0.5 cycles/degree (*r* = .24). These latter correlations were not in the expected direction however. We anticipated that better contrast sensitivity (a more negative number) would be associated with a longer (not shorter) VID range, thus producing a negative rather than a positive correlation.

An additional significant correlation was found between the number-of-times-killed score and low contrast acuity with glare threshold mean GLAR-TM (*r* = -.22). However, the negative correlations indicated that better performers on this vision test were killed more often in simulated ACM! Again, age or experience may be affecting these correlations since the vision of older pilots may be more susceptible to the effects of glare. Older pilots, however, may use their experience to an advantage in ACM.

In summary, the results indicated that vision tests scores were related to components of ACM performance that were associated with highly vision dependent tasks. This should not be too surprising. The vision test battery was specifically designed to identify critical visual functions predictive of success in ACM, particularly, the range of initial target detection and identification. Thus, we did not expect the NAMRL vision tests to correlate with all aspects of ACM performance. For example, VID kills are accomplished on the initial pass where performance may reasonably be dependent on the range of initial visual acquisition. In contrast, an engaged kill occurs during the portion of ACM when each pilot tries to maneuver behind the other to fire. Engaged kills may require fewer visual skills and are affected comparatively more by psychomotor skills, the skills of the adversary, and particularly tactics associated with weapon system employment.

Table 8 presents correlations between ACM performance criteria and measures of experience--specifically, age, jet hours, total jet hours, TACTS hours, and total ACM flight hours. Of 45 correlations, 8 (18%) were significant at the .05 or .01 level of confidence. The one criterion consistently related to age or flight experience measures was the mean time-to-first-kill score, which produced significant correlations with age (*r* = - = -.22). In each case, greater age or more flight experience was associated with shorter mean time-to-first-kill scores (better performance). Visual ID performance was significantly related to jet hours (*r* = .24) and total ACM flight hours (*r* = .21). The total number of ACM flight hours was the only experience measure significantly related to the OAG (*r* = .23) and the kill difference score (*r* = .21). There were no significant correlations between age or flight experience measures and situational awareness, visual tally range, VID range, number of times killed, and the number of engaged kills. Further, only one experience measure, total ACM hours, was significantly related to the OAG, VID kills, and the kill-difference score. These correlations were generally quite small, accounting for only 4 or 5% of the variance with each criterion.

TABLE 8. Pearson Correlations for Experience Measures and ACM Performance (n varies).

ACM performance criteria	Age n=89	Jet hours n=88	Total hours n=88	TACTS hours n=85	Total ACM n=85
OAG	.10	.20	.17	.13	.23*
Situational awareness	.04	.08	.06	-.02	.08
Mean time-to-first-kill	-.34**	-.36**	-.32**	-.22*	-.16
Visual tally mean range	.05	-.01	-.04	.06	-.01
Visual ID mean range	-.02	-.06	-.11	-.08	.06
Number of times killed	-.06	-.05	.03	.14	-.02
VID kills	.09	.24*	.17	.16	.21*
Engaged kills	.13	.10	.13	-.02	.13
Kill-difference score	.06	.10	.07	.04	.21*

* $p < .05$
** $p < .01$

Our results indicate that ACM experience influences ACM performance, especially in achieving VID kills and improved time-to-first-kill scores. On the other hand, these results also suggest that experience is not related to situational awareness, visual tally, VID range, engaged kill, or being killed criteria. Future evaluations of ACM TACTS performance should include experience factors similar to those examined here to better understand the relation of age and experience to TACTS ACM performance.

DISCUSSION AND CONCLUSIONS

Study I. Multitask Test and Marine Pilot ACM Performance.

This evaluation was conducted to test the feasibility of predicting ACM performance with perceptual motor and cognitive multitask tests. Eighteen F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed automated performance-based tests.

Initial analyses indicated that the overall ACM grade (OAG) associated with the VF-43 adversary squadron's evaluation of ACM performance of F-4 pilots was significantly and positively related to offensive maneuvering, situational awareness, and mutual support measures. In addition, the OAG was significantly but negatively related to the objective TACTS measures, mean time-to-first-kill, adversary squadron missile shots, and the number of times a pilot was killed in the simulated exercises. The negative correlations indicated that a higher ACM grade was associated with shorter times to first kill, fewer adversary squadron missile shots taken, and fewer times being "killed" in the simulated exercises. A multiple regression analysis indicated that four of these measures, situational awareness, offensive maneuvering, number of times killed, and mutual support, could reliably predict the OAG criterion.

The OAG and the four criterion measures were then correlated with the single- and multitask cognitive and psychomotor test performance of the F-4 pilots. A derived composite kill-difference score, based on the total number of ACM kills less the number of times a pilot was killed in the TACTS simulated

exercises, was included in the correlational analyses as well.

A DLT measure obtained during multitask performance was significantly related to offensive maneuvering and the kill-difference composite score. A DLT multitask test measure based on a slightly different scoring procedure was significantly related to the OAG and the offensive maneuvering score.

Conclusion: Multitask tests reliably predicted ACM performance for a small sample of F-4 pilots.

Unresolved, however, was a series of important questions concerning the Fleet Fighter ACM Readiness Program evaluation process: Are the resulting grades reliable for Navy pilots flying contemporary F-14 aircraft? What is the relation of the VF-43 grading process to more objective TACTS ACM performance measures (i.e., total number of kills, VID range, VID kills, and engaged kills)? And which ACM measures are most predictive of ACM performance?

Study II: Fleet Fighter ACM Readiness Program Grades as Criteria.

Objectives of Study II were to identify criteria for the validation of tests designed to predict ACM performance and estimate the reliability of readiness grades used to assess Navy pilot ACM proficiency.

An examination of subjective and objective measures of the TACTS ACM competitive exercise performance of 125 naval aviators participating in Fleet Fighter ACM Readiness Program Exercises at NAS Oceana, indicated that the overall grade (OAG) can be reliably predicted by a relatively few measures. These were a kill-difference composite score (resulting from adding the number of TACTS VID and engaged kills and then subtracting the number of times a pilot was killed during TACTS competitive exercises), situational awareness, energy management, and mutual support measures.

A separate correlational analysis examined the reliability of the Fleet Fighter ACM grading process. Subjects were randomly divided into two groups and the correlations between the various performance measures and OAG were examined. The resulting $\underline{r}$ values were highly similar. In addition, correlation values were computed for eight different pilot subsamples based on a total group prediction model of OAG and the OAG actually obtained. The average of the eight correlation values was .88. These results suggest that the Fleet Fighter Readiness grading process is reliable. Regardless of the subject sampling procedure, the model for predicting OAG provided highly similar results. Apparently the grading of ACM performance by different VF-43 adversary pilots was consistent.

Conclusion: Fleet Fighter ACM Readiness program grades are reliable and suitable criteria for validating tests designed to predict F-14 pilot ACM performance.

Study III: The Relation of Vision Test, Experience, and ACM Performance.

The purpose of Study III was to examine the relation between experimental NAMRL vision tests and ACM performance measures identified in Study II. Contrast sensitivity tests were significantly correlated with a mean time-to-first-kill score, and visual acuity and accomodative flexibility

tests were significantly correlated with TACTS visual tally-ho and VID range criteria. Many of the vision and TACTS performance correlations seemed to be affected by an age or experience factor. For example, pilots who obtained shorter mean times-to-first-kills had poorer visual contrast sensitivity and acuity, but also were more experienced (and older) and apparently used this experience to achieve faster kills in ACM.

Conclusion: The NAMRL vision test scores were related more to the highly vision-dependent components of ACM, e.g., visual tally range and VID range.

Additional correlations were computed between ACM performance criteria and measures of experience--specifically, age, jet hours, total jet hours, TACTS hours and total ACM flight hours. The one criterion consistently related to age or flight experience measures was the mean time-to-first-kill score, producing significant correlations with age, jet hours, total flight hours, and TACTS hours. In each case, greater age or more flight experience was associated with shorter mean time-to-first-kill scores (better performance). Visual ID kill performance was significantly related to jet hours and total ACM flight hours. Total ACM flight hours was significantly related to the OAG, VID kills, and the kill-difference score. There were no significant correlations between age or flight experience measures and situational awareness, visual tally range, VID range, number of times killed and the number of engaged kills.

Conclusion: Experience in ACM influences performance, especially in achieving VID kills and improved time-to-first-kill scores. Experience in ACM was <u>not</u> related to situational awareness, visual tally, VID range, engaged kill, and being killed criteria.

RECOMMENDATIONS

Results (Study I) demonstrated the feasibility of using automated, synthetic, cognitive, perceptual, and multitask tests to predict TACTS F-4 pilot ACM proficiency and indicated (Study II) that Fleet Fighter ACM Readiness Program grades are reliable criteria for validating tests designed to predict ACM performance.

To achieve the goal of validating tests to aid in aircrew selection and assignment decisions, the following research is needed.

1. Synthetic cognitive, perceptual, and multitask tests should be administered to a suitable sample of F-14 pilots performing in Fleet Fighter Readiness Evaluations to replicate initial test results.

2. Pilot experience data should be included in the above effort to study the relation of age and experience to TACTS ACM performance.

3. An analysis of additional OAG data would be useful in assessing the reliability and validity of mathematical models to predict Fleet Fighter ACM grades.

The successful validation of synthetic tests to predict ACM performance would be valuable for improving the quality and capabilities of fighter aircrew through their initial selection and subsequent assignment to training pipelines and aircraft.

REFERENCES

1. Ciavarelli, A, Brictson, C. A., and Young, P. A., Development and Application of Performance Criteria and Aircrew Assessment Methods for the Air Combat Maneuvering Range (ACMR), NAMRL Special Report 79-5, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1979.

2. Bale, R. M., Rickus, G. M., and Ambler, R. K., Replacement Air Group Performance as a Criterion for Aviation Training, NAMRL 1126, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

3. Brictson, C. A., Burger, W. J. and Gallagher, T. J., "Prediction of Pilot Performance during Initial Carrier Landing Qualifications." Aerospace Medicine, Vol. 43, pp. 483-487, 1972.

4. Rickus, G. M. and Berkshire, J. R., Development of an Aviation Combat Criterion, NAMI-1047, Naval Aerospace Medical Institute, Pensacola, FL, 1968.

5. Shannon, R. H. and Waag, W. L., The Prediction of Pilot Performance in the F-4 Aircraft, NAMRL-1186, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

6. Shannon, R. H. and Waag, W. L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, NAMRL-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

7. Shannon, R. H., Waag, W. L., and Ferguson, J. C. A New Approach to Criterion Development in the Replacement Air Group, NAMRL-1158, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

8. SAS User's Guide: Statistics (5th ed.), SAS Institute, Inc. Cary, NC, 1985.

9. Commander Fighter Wing One Instruction 3710.3E, Naval Air Station Oceana, Virginia Beach, VA, 11 Feb 1985.

10. Pitts, D. G., "The Effects of Aging on Selected Visual Functions: Dark Adaptation, Visual Acuity, Stereopsis, and Brightness Contrast." In R. Sekuler, D. Kline, and K. Dismukes. (Eds.), Aging and Human Visual Function, Alan R. Liss, Inc., NY, 1982, pp. 131-160.

11. Sekuler, R. and Owsley, C., "The Spatial Vision of Older Humans." In R. Sekuler, D. Kline, and K. Dismukes. (Eds.), Aging and Human Visual Function, Alan R. Liss, Inc., NY, 1982, pp. 185-202.

12. Monaco, W. A. and Hamilton, P. V., "Visual Capabilities Related to Fighter Aircrew Performance in the F-14 and Adversary Aircraft." Advisory Group for Aerospace Research and Development Conference Proceedings No. 396, pp. 38-1 to 38-9, December 1985.

13. Morris, A., Hamilton, P. V., Morey, W. A., and Briggs, R. P., "Vision Test Battery Threshold and Response Time as Predictors of Air-to-air Visual Target Acquisition in F-14 and Adversary Aircraft." Advisory Group for Aerospace Research and Development Conference Proceedings No. 396, pp. 39-1 to 39-8, December 1985.

APPENDIX A

APPENDIX A

The Tactical Aircrew Combat Training System (TACTS) is one of the most technologically sophisticated training systems in existence. The system is a computer based tracking and data communication network that enables ACM training and simulated weapons firing of actual aircraft engaged in ACM in real time. The TACTS system provides military managers with visual flight dynamics, weapons system status, and weapons firing information. All data (including the visual representation of aircraft) are recorded on magnetic tape for use in the debriefing of both adversary and fighter aircrews. The need for such a training system became apparent during the Viet Nam conflict in which pilots often failed to recognize when they were in a correct firing envelope for the missile weaponry of that day. The TACTS system enables the employment of both rear-aspect and forward-aspect missile weaponry and serves as a means of evaluating the tactical use of both missile and aircraft weapon systems in simulated air combat.

APPENDIX B

APPENDIX B

ABBREVIATED ACM PERFORMANCE MEASURE DEFINITIONS

01. OVERALL ACM GRADE - a composite of 19 subjective measures - (see page 4 of text, Figure 1).
02. ENVIRONMENT - use of weather conditions to gain an advantage in ACM.
03. START/VID START - position at start of engagement when the fighter and adversary aircraft merge*.
04. FIRST MOVE - a positioning advantage the fighter tries to obtain just before the merge*.
05. AGGRESSIVENESS - how aggressively the fighter employs his aircraft weapon systems.
06. OFFENSIVE MANEUVERING - fighter's ability to optimize offensive position and achieve missile shots.
07. DEFENSIVE MANEUVERING - fighter's ability to maneuver while defensive and avoid being shot.
08. KEEPING SIGHT - awareness of position of wingmen and adversary aircraft.
09. ENERGY MANAGEMENT - optimizing airspeed while maneuvering.
10. MENTAL PLOT - fighter's mental picture of aircraft positioning while engaged.
11. SITUATIONAL AWARENESS - the total of ACM performance.
12. BUGOUT - technique used to disengage from ACM and arrive at a safe area.
13. WEAPON EMPLOYMENT - radar use in intercept and use of weapons while engaged in ACM.
14. VID TECHNIQUE - appropriate use of radar in the intercept.
15. VID COMMUNICATION - fighter-to-fighter and fighter-to-ground control, radar intercept communications.
16. UHF COMMUNICATION - fighter to fighter communication while engaged.
17. GAME PLAN - execution of tactical engagement plan.
18. MUTUAL SUPPORT - fighter's ability to protect and support wingmen.
19. DEBRIEF - participation in the fighter/adversary debriefing.
20. RECONSTRUCTION - ability to remember and reconstruct the ACM fight.
21. NUMBER OF KILLS - combination of measures 23 and 24.
22. NUMBER OF MISSILES LAUNCHED - Self explanatory.
23. NUMBER OF VID KILLS - pre-merge* kills. (These are made prior to actual ACM, usually with forward aspect missiles.)
24. NUMBER OF ENGAGED KILLS - post-merge* kills. (Those made during actual ACM, usually with heat seeking missiles.)
25. MEAN TIME FIRST KILL - calculated from 10-mile separation point of fighter and adversary aircraft.
26. NUMBER OF RADAR LOCKS - Self explanatory.
27. RADAR LOCK MEAN RANGE - mean range at which radar lock obtained.
28. VISUAL TALLY-HO MEAN RANGE- mean range of initial sighting of adversary aircraft during intercepts.
29. VID MEAN RANGE - mean range of adversary aircraft identifications, i.e., "A-4."
30. NUMBER OF TIMES KILLED - Self explanatory.
31. KILL DIFFERENCE SCORE - measure 21 minus 30 - a composite score.
32. ENGAGED KILL DIFFERENCE SCORE - measure 24 minus 30 - a composite score.

* Merge point: The point at which the fighter and adversary aircraft first pass during the intercept.

APPENDIX C

TABLE C-1. Correlation Matrix of TACTS Subjective and Objec

	1	2	3	4	5	6	7	8	9	10	11
01. Overall ACM grade	- 1										
02. Environment	+.002	- 2									
03. Start/VID start	+.404	-.106	- 3								
04. First move	+.119	-.012	-.087	- 4							
05. Aggressiveness	+.449	-.040	+.148	+.074	- 5						
06. Offensive man.	+.529	-.012	+.121	-.066	+.300	- 6					
07. Defensive man.	+.393	-.105	+.277	-.068	+.280	+.141	- 7				
08. Keeping sight	+.350	-.038	+.110	-.019	-.006	+.107	-.075	- 8			
09. Energy management	+.371	+.006	+.194	+.171	-.026	+.248	+.103	+.057	- 9		
10. Mental plot	-.108	-.061	+.048	-.051	-.055	-.052	+.084	-.046	+.210	-10	
11. Situational aware	+.697	+.014	+.209	+.004	+.319	+.298	+.235	+.256	+.056	-.155	-
12. Bugout	+.112	-.116	+.110	-.156	-.181	-.065	-.016	+.064	-.041	-.054	+.0
13. Weapon employment	+.298	-.085	+.102	-.093	-.160	+.104	+.031	+.202	+.045	-.139	+.0
14. VID technique	+.302	-.124	+.132	+.182	+.202	-.032	+.139	+.134	+.153	+.066	+.
15. VID comm.	-.078	+.123	-.154	+.013	-.039	-.069	-.090	+.099	-.019	+.038	-.
16. UHF comm.	+.115	+.062	+.214	-.105	+.145	+.028	+.155	-.009	-.008	-.025	+.
17. Game plan	+.322	+.059	-.018	+.016	+.159	+.147	-.123	+.053	+.145	-.075	+.
18. Mutual support	+.435	+.142	-.066	+.243	+.191	+.092	+.108	-.093	+.034	-.330	+.
19. Debrief	+.110	-.036	+.037	-.008	+.290	+.026	-.066	-.105	+.009	-.053	+.
20. Reconstruction	+.162	+.037	-.123	+.235	+.138	-.018	+.068	-.101	+.066	+.092	+.
21. Total TACTS kills	+.648	+.131	+.253	-.024	+.265	+.472	+.132	+.278	+.242	-.113	+.
22. Missiles launched	+.582	+.148	+.283	+.086	+.284	+.392	+.111	+.255	+.271	-.060	+.
23. No.VID kills	+.391	+.078	+.181	-.145	+.132	+.304	+.072	+.135	+.201	-.021	+.
24. No.engaged kills	+.572	+.116	+.201	+.086	+.257	+.401	+.122	+.271	+.169	-.137	+.
25. Time first kill	-.024	+.089	-.058	+.004	+.003	+.066	-.121	+.166	+.052	+.001	-.
26. No.radar locks	+.235	+.103	+.171	-.022	-.101	+.194	+.074	+.227	+.197	-.064	+.
27. Radar lock range	-.040	+.030	+.001	+.080	-.182	-.165	-.074	+.139	+.002	-.161	-.
28. Eyetally x range	+.118	+.134	+.180	-.222	-.054	+.170	+.125	+.053	+.055	-.052	+.
29. Eye ID x range	+.156	+.044	+.185	-.051	-.015	+.189	+.021	+.119	+.133	+.015	+.
30. No.times killed	-.514	+.032	+.199	-.075	-.210	-.075	-.282	-.192	-.219	-.184	-.
31. Kill dif.score	+.757	+.107	+.285	+.005	+.314	+.449	+.220	+.312	+.292	-.037	+.
32. Engaged killed difference score	+.704	+.084	+.262	+.107	+.305	+.351	+.224	+.312	+.241	-.039	+.

.05 = +.176
.01 = +.230

1 of 2

d Objective Performance Measures.

)																							
.0	-10																						
36	-.155	-11																					
1	-.054	+.025	-12																				
5	-.139	+.041	-.012	-13																			
53	+.066	+.206	+.105	-.192	-14																		
19	+.038	-.049	-.132	-.115	-.012	-15																	
8	-.025	+.147	-.177	-.296	+.192	+.064	-16																
5	-.075	+.195	-.085	+.035	-.009	+.209	+.035	-17															
34	-.330	+.231	+.002	+.127	-.036	-.181	-.086	+.067	-18														
9	-.053	+.083	+.035	-.077	+.007	+.042	+.176	+.140	+.049	-19													
36	+.092	+.045	-.062	-.024	+.124	+.060	-.007	+.242	+.182	+.074	-20												
2	-.113	+.411	-.019	+.323	+.131	-.012	+.080	+.290	+.218	+.019	-.024	-21											
71	-.060	+.385	-.068	+.179	+.100	+.079	+.061	+.286	+.196	+.018	+.035	+.827	-22										
1	-.021	+.324	-.008	+.223	+.062	+.055	-.003	+.251	-.007	-.009	+.010	+.689	+.571	-23									
59	-.137	+.301	-.020	+.264	+.130	-.061	+.112	+.195	+.303	+.034	-.040	+.814	+.672	+.141	-24								
52	+.001	-.115	+.051	-.046	-.062	-.055	-.135	+.009	-.012	+.082	-.148	+.006	+.110	-.428	+.354	-25							
37	-.064	+.125	+.020	+.208	+.002	+.178	+.147	+.039	-.114	+.102	+.032	+.409	+.397	+.422	+.221	-.119	-26						
2	-.161	-.067	+.045	+.156	+.095	+.014	+.214	-.111	-.063	+.017	+.112	-.104	-.208	-.118	-.048	-.006	+.249	-27					
55	-.052	+.121	+.073	+.131	-.112	+.108	+.020	+.042	-.135	-.065	-.184	+.303	+.323	+.449	+.054	-.189	+.434	+.050	-28				
33	+.015	+.150	-.101	+.119	-.087	+.203	+.085	+.138	-.162	-.010	-.159	+.405	+.420	+.484	+.165	-.141	+.495	-.070	+.676	-29			
19	-.184	-.390	-.227	-.073	-.117	+.132	-.023	-.108	-.292	-.033	-.128	-.117	-.118	-.024	-.140	-.114	+.057	+.004	+.103	+.067	-30		
92	-.037	+.500	+.051	+.322	+.157	-.059	+.080	+.299	+.300	+.030	+.025	+.933	+.775	+.620	+.777	+.048	+.344	-.094	+.228	+.333	-.457	-31	
41	-.039	+.425	+.086	+.241	+.180	-.111	+.105	+.224	+.381	+.044	+.024	+.733	+.617	+.131	+.896	+.341	+.140	-.047	+.003	+.113	-.556	+.851	-

2 of 2

APPENDIX D

APPENDIX D

TABLE D-1. Subjective Measure Forward Selection, Analysis of Variance, Coefficient, *F* Values, and Model Summary Statistics.

Stepwise regression procedure for dependent variable (overall ACM grade)

Step 6, Variable V8 entered — Multiple *R* = .91; *R* Square = 0.83; Adjusted *R* Square (shrinkage) = *R* = .81

	df	Sum of Squares	Mean Square	*F*	*p*
Regression	6	0.2658	0.0443	98.57	0.0001
Error	118	0.0530	0.0004		
Total	124	0.3189			

	B Value	Standard Error	Type 11 SS	*F*	*p*
Intercept	1.1111				
V3	0.0760	0.0124	0.0169	37.62	0.0001
V6	0.0732	0.0111	0.0195	43.49	0.0001
V8	0.0747	0.0143	0.0123	27.37	0.0001
V9	0.0731	0.0136	0.0129	28.70	0.0001
V11	0.0869	0.0087	0.0445	99.08	0.0001
V18	0.0669	0.0077	0.0336	74.66	0.0001

Summary of stepwise regression procedure for dependent variable V1 - OAG

Step	Variable Entered	Number In	Partial *r2*	Model *r2*
1	V11 Situational awareness	1	0.4862	0.4862
2	V6 Offensive maneuvering	2	0.1132	0.5995
3	V18 Mutual support	3	0.0747	0.6742
4	V3 Start/VID start	4	0.0777	0.7519
5	V9 Energy management	5	0.0432	0.7951
6	V8 Keeping sight/lookout	6	0.0386	0.8337

TABLE D-2. Objective Measure Forward Selection, Analysis of Variance, Coefficients, F Values, and Model Summary Statistics.

Stepwise regression procedures for dependent variable (overall ACM grade)

Step 4, Variable V25 entered

Multiple R = .79

R Square = 0.63

Adjusted R Square (shrinkage) = R = .61

	df	Sum of Squares	Mean Square	F	p
Regression	4	0.2008	0.0502	51.02	0.0001
Error	120	0.1181	0.0010		
Total	124	0.3189			

	B Value	Standard Error	Type 11 SS	F	p
Intercept	1.9973				
V23	0.0075	0.0020	0.0134	13.62	0.0003
V24	0.0131	0.0016	0.0671	68.17	0.0001
V25	-0.0003	0.0001	0.0044	4.44	0.0371
V30	-0.0205	0.0026	0.0624	63.40	0.0001

Bounds on condition number: 1.5855, 42.7055

Summary of stepwise regression procedure for dependent variable V1 - OAG

Step	Variable Entered	Number In	Partial $r2$	Model $r2$
1	V24 Engaged kills	1	0.3273	0.3273
2	V30 Number times killed	2	0.1919	0.5192
3	V23 VID kills	3	0.0968	0.6160
4	V25 Mean time to first kill	4	0.0137	0.6297

TABLE D-3. Subjective and Objective Measure Forward Selection, Analysis of Variance, coefficients, F Values, and Model Summary Statistics.

Stepwise regression procedure for dependent variable (overall ACM grade)
Step 4, Variable V18 entered
Multiple R = .89
R Square = 0.78
Adjusted R Square (shrinkage = R = .78

	df	Sum of Squares	Mean Square	F	p
Regression	4	0.2503	0.0626	109.39	0.0001
Error	120	0.0686	0.0006		
Total	124	0.3189			

	B Value	Standard Error	Type II SS	F	p
Intercept	1.5826				
V11	0.0880	0.0101	0.0432	75.61	0.0001
V9	0.0757	0.0154	0.0138	24.12	0.0001
V18	0.0400	0.0088	0.0119	20.78	0.0001
V31	0.0067	0.0008	0.0358	62.59	0.0001

Bounds on condition number: 1.5474, 41.0336

Summary of stepwise regression procedure for dependent variable V1 - OAG

Step	Variable Entered	Number In	Partial $r2$	Model $r2$
1	V31 Kill difference score	1	0.5724	0.5724
2	V11 Situational awareness	2	0.1356	0.7080
3	V9 Energy management	3	0.0395	0.7475
4	V18 Mutual support	4	0.0373	0.7848

☆U.S. GOVERNMENT PRINTING OFFICE: 1988 - 530-012-65453

Other Related NAMRL Publications

Griffin, G.R. and McBride, D.K., Multitask Performance: Predicting Success in Naval Aviation Primary Flight Training, NAMRL-1316, Naval Aerospace Medical Research Laboratory, Pensacola, FL, March 1986. (AD# A168 246)*

Griffin, G.R. and Williams, C.E., "The Effects of Different Levels of Task Complexity on three Vocal Measures." Aviation Space, and Environmental Medicine, Vol. 58, No. 12, pp. 1165-1170, December 1987.

* This publication is available from DTIC, Cameron Station, Alexandria, VA 22314 (Phone: (C) 202/274-7633 or (A) 284-7633). Use the AD number to request reports. The report listed without an AD number may be requested from the NAMRL author or Code 00A4.

DEPARTMENT OF THE NAVY

COMMANDING OFFICER
NAVAL AEROSPACE MEDICAL
RESEARCH LABORATORY
NAVAL AIR STATION
PENSACOLA, FL 32508-3700

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

U.S. MAIL

POSTAGE AND FEES PAID
DEPARTMENT OF THE NAVY
DOD 316